DOCUMENT RESUME

ED 074 154 TM 002 519

AUTHOR Knipe, Walter H.; Krahmer, Edward F.
TITLE An Application of Criterion Referenced Testing.
PUB DATE 27 Feb 73
NOTE 19p.; Paper presented at annual meeting of the American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Computer Assisted Instruction; *Criterion Referenced Tests; Elementary Grades; *Mathematics Instruction; *Performance Contracts; *Program Evaluation; Speeches; Teacher Attitudes; Technical Reports; Testing Programs

ABSTRACT

Criterion referenced testing has received considerable theoretical, but only limited practical, application. Grand Forks School District has developed mathematics criterion referenced tests for grades three to nine. The tests are keyed to a hierarchical set of approximately 50 performance objectives and 40 individualized contracts per grade level. These tests were administered on a pre-post basis during 1971-72. This study was designed to consider the following three concerns: (1) adoption experiences when using criterion referenced testing, (2) research conclusions as a byproduct of this testing, and (3) attitudes of classroom teachers to this method as compared to nationally formed tests. Data for the second concern indicate different orders and grade levels at which students learn specific skills in various schools. (Author)

FILMED FROM BEST AVAILABLE COPY

ED 074154

An Application of Criterion Referenced Testing

Presented at

American Education Research Association Convention
February 27, 1973

TM 002 519

Walter H. Knipe
Dr. Edward F. Krahmer

Not for publication, all rights reserved

AN APPLICATION OF CRITERION REFERENCED TESTING

Walt Knipe, Director, Learner Orientation to Technology Project
Grand Forks (North Dakota) School District

Edward Krahmer, Director, Resource Information Center
Grand Forks, North Dakota

Introduction

Criterion referenced testing is becoming a household byword; it is often referred to as the "answer" to education's test and measurement problems. This is obviously a very ambitious responsibility to assign to a relatively recent concept. Behavioral objectives are another example of something once considered to be a panacea for education, which is now being seen in a more realistic light. This paper attempts to consider some of the more realistic possibilities for criterion referenced testing in a local school setting. The results for the most part are preliminary since the project and research study from which the data contained in this paper were obtained is ongoing.

Objectives

Since the intent of this paper is to present some of the research byproducts of an application of criterion referenced testing to a medium sized school district, the objectives are as follows:

1. What are the adoption experiences of a school district with regard to criterion referenced testing?

2. What are the primary research conclusions resulting as byproducts from the criterion referenced experience of a school district?

3. After two years experience with criterion referenced testing, what are the attitudes of the classroom teachers to this method of testing in comparison to traditional national norm-referenced tests?

Part 1 Adoption Experiences

The experience of the Grand Forks (North Dakota) School District, a district having an enrollment of approximately 12,000, with regard to criterion referenced testing dates back to at least 1967-68 when the first attempts were made to develop individualized teacher/learner contracts. The summer of 1969, the school district employed over a hundred teachers (using local monies) to specify in detail the performance objectives for k-12 in most subject areas and to begin the development of a comprehensive set of individualized teacher/learner contracts as one instructional method whereby students could meet these objectives. This ambitious undertaking, now known regionally as the Grand Forks Learning System, resulted in over 3,500 contracts the first year and over 4,000 by the present time.

Obviously, not all grade levels or subject areas received the same degree of attention. Mathematics, however, being quite amenable to an individualized

contracts approach, probably received the most consistent and extensive treatment by grade levels. An average of something in excess of 50 objectives per grade level developed on a hierarchical skills basis with approximately 40 contracts per grade to fulfill these objectives resulted.

The district administration, early recognizing the necessity of a management system for maintaining the Learning System, encouraged the mathematics department of one junior high school to seek Title 111 ESEA support from the State of North Dakota to establish a Computer assisted and computer managed instructional model (CAI and CMI) paralleling the Learning System. This project, Learner Orientation to Technology (LOTT), now in its third year of operation, immediately focused on the area of testing as the primary pressing problem with regard to utilizing CAI, CMI, and individualized contracts.

During the summer of 1971 a number of mathematics teachers were employed to develop test items to be used in measuring achievement of the performance objectives From one to three items were developed per objective; the number of items depending upon whether or not the objective provided a variety of options for its achievement. Criterion referenced tests of approximately 120 items resulted for each grade level from 3 through 9. Since no overlap in objectives occurs from one grade level to the next, no overlap in test items occurred. Overlap in terms of mathematics skills, however, does occur from grade level to grade level, and, as such, is measured. The student is expected to correctly answer 1 of 1, 1 of 2, or 2 of 3 items as evidence of successful obtainment of each objective. Each test item had five possible responses.

Tests were printed for each grade level of each school by the school district print shop, and administered in early October of 1971. Each teacher was provided with an instruction manual which explained how the tests should be administered and some suggested uses for the results.

Students marked their responses on a mark sense computer card, which contained their name and student I, D. number.

Some of the reactions to criterion referenced testing after a two year period of time will be found in the conclusion of this paper.

Part 11 Criterion-Referenced Test Research Conclusions

The locally developed criterion referenced tests have been administered twice on a district-wide basis to all students in grades 3 through 8. Both administrations were during 1971-72; the Fall administration for baseline and placement purposes and the Spring administration for follow-up and for placement in the Fall of 1972-73.

At the present time the possible side benefits which can result from performing statistical analysis on the data obtained from administration of criterion-referenced tests are only just being realized by district personnel. These side benefits are being identified as a result of the observation that there are obvious differences between schools and even amoung classes within schools as to the emphasis placed on the order and sequence of student fulfillment of performance objectives.

While no sophisticated statistical analysis has yet been performed on the data, th the results presented should justify the conclusions that:

a. different learning patterns exist from school to school

b. even within a school, different learning patterns exist for different classes.

To demonstrate the first conclusion, 50 fifth grade pupils were randomly selected from each of three elementary schools and the same number of eighth graders from three junior high schools. A pictorial demonstration of the different learning patterns was developed by scoring the criterion referenced test from the Spring administration and plotting correct responses (using as asterisk) in columns representing performance objectives. The following six pages indicate the results obtained.

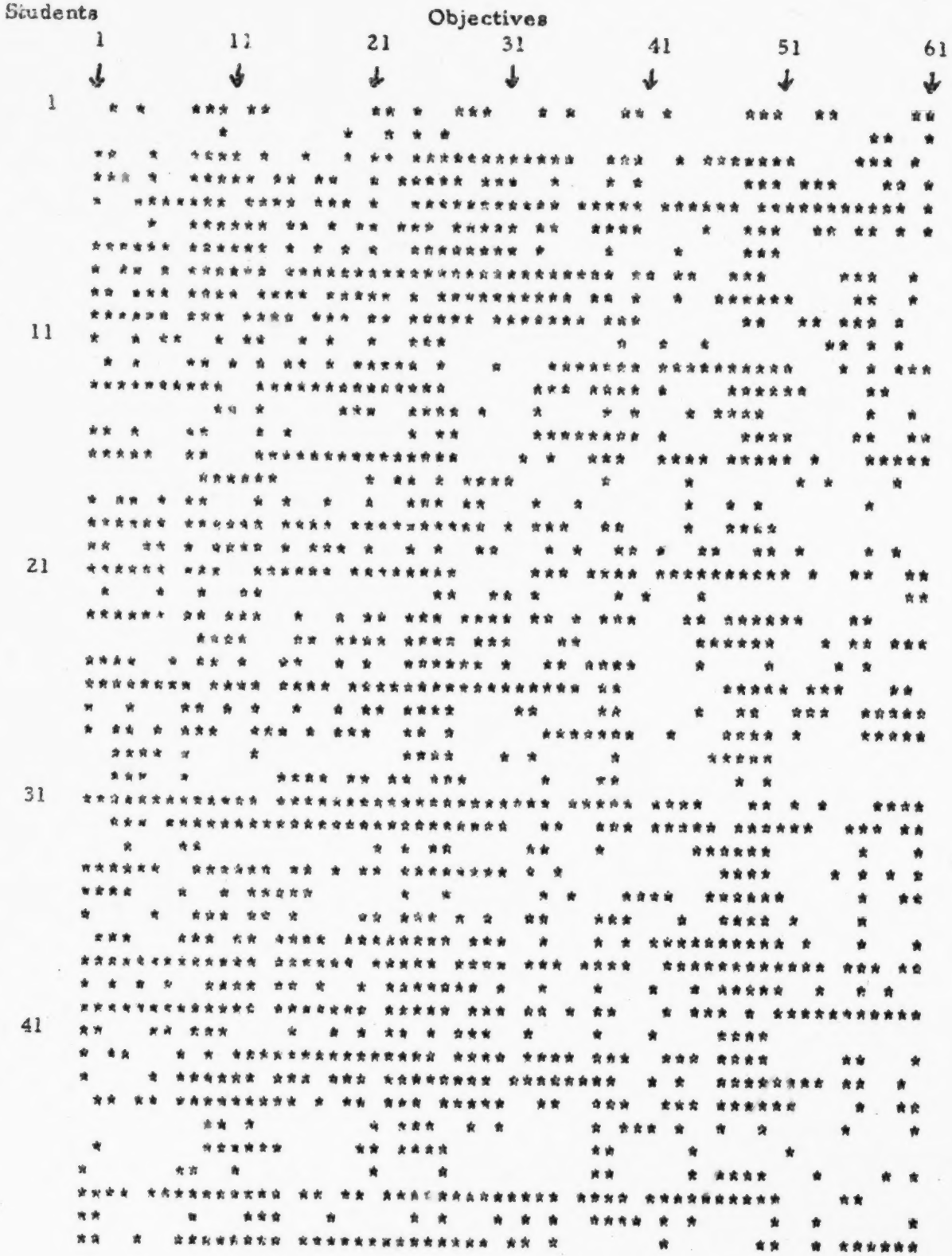
Looking at one example, Grade 8; it will be noted in the column headed objective 1 that the following data resulted;

School	Number of Correct Responses	percent
A	34	68.0
B	23	46.0
C	27	54.0

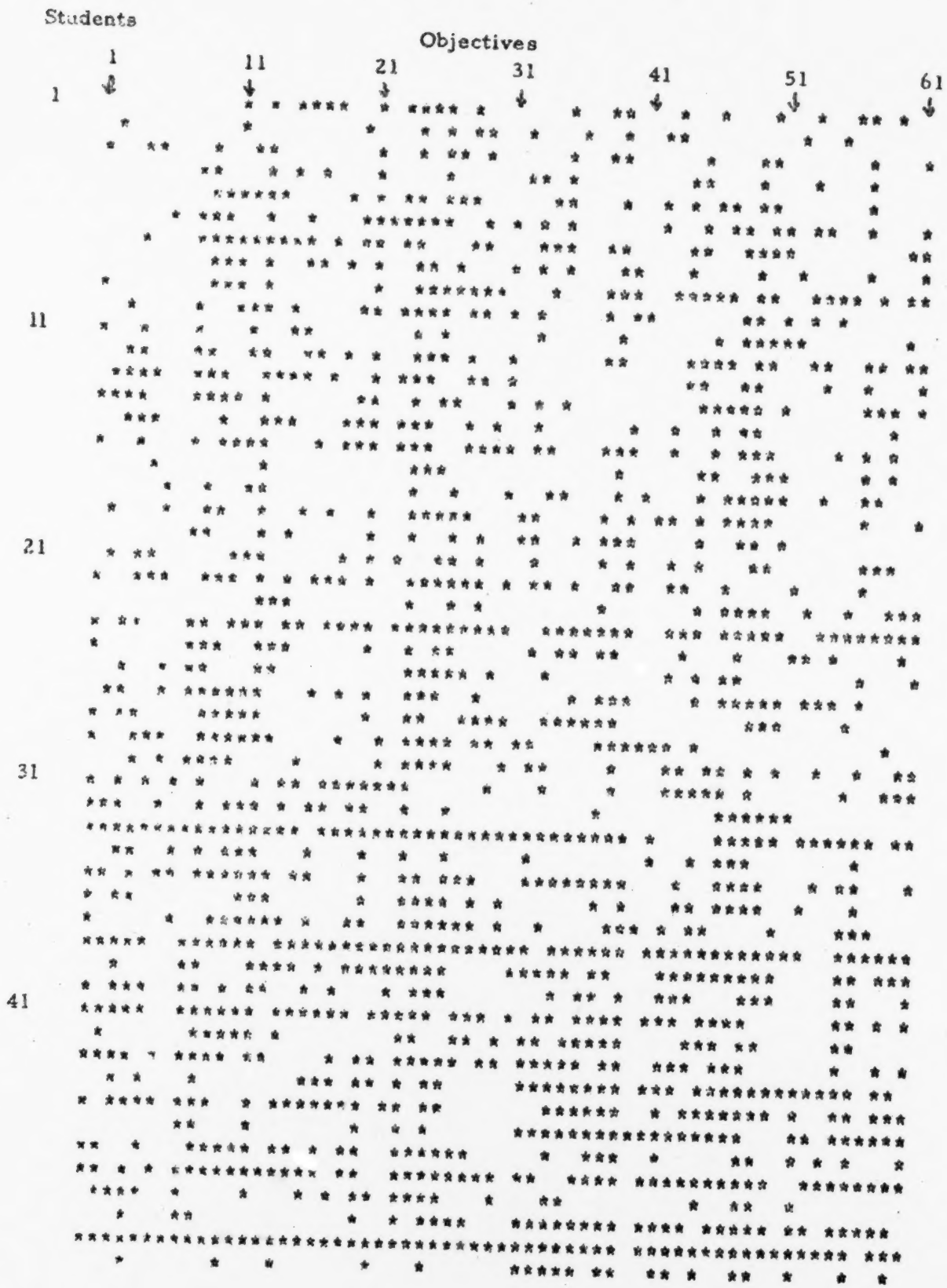
Looking at the list of performance objectives measured by this test, it will be noted that objective 1 refers other Bases to Base 10. (See Appendix A at the conclusion of this report for a list of objectives for grades 5 and 8). Thus two thirds of the students in School A had met this objective by the end of eighth grade, while less than half had met it in School B, a considerable difference.

To demonstrate the second conclusion, the same pictorial method was used, only here the students were randomly selected from the eighth grade classes of each teacher in one school. The following figures provide the results for two example teachers. Looking at the last objective, Relative error, students in Teacher A's classes met this objective in 30 cases (60%) while students in Teacher B's classes met the objective 41 times (82%).

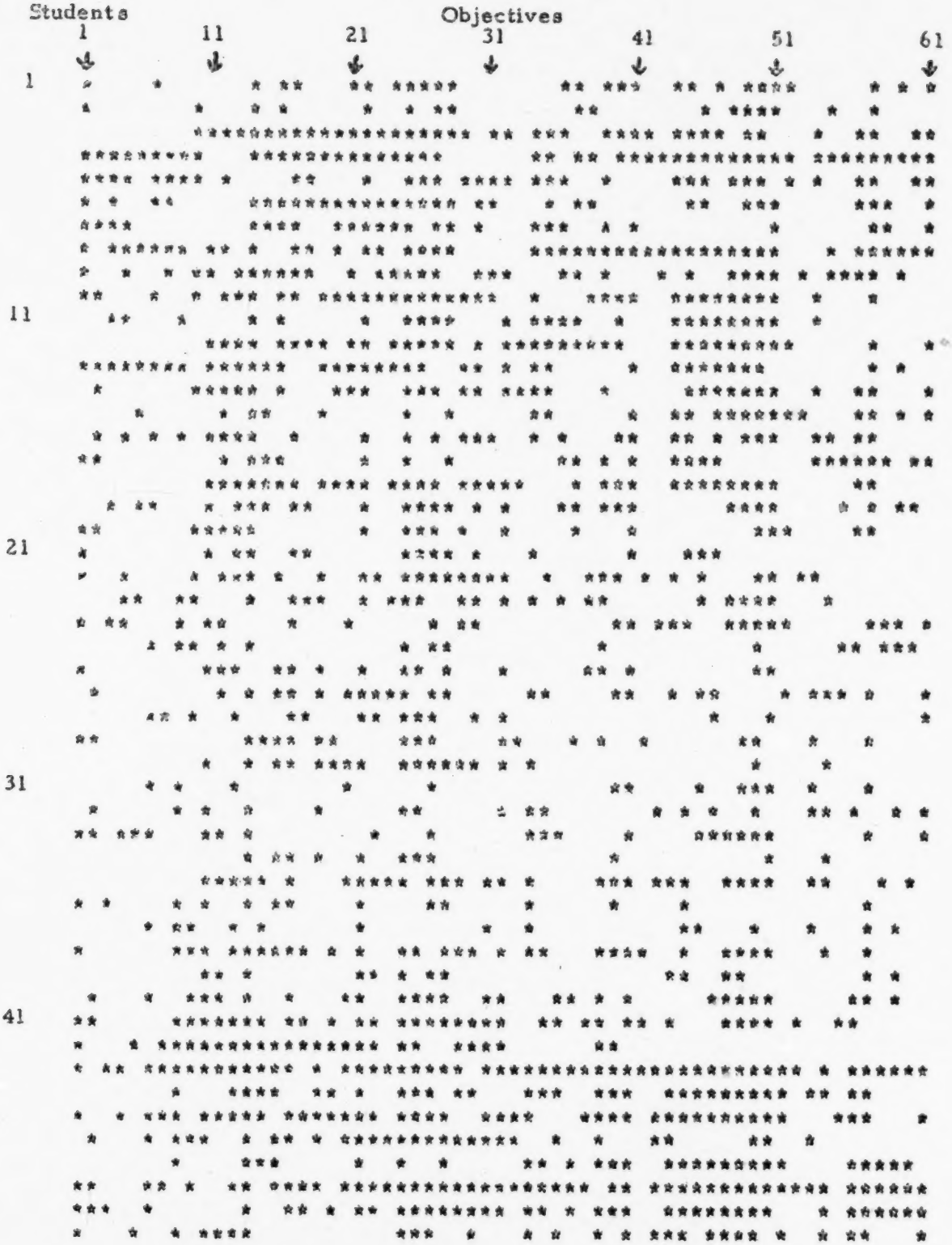
School A-Grade 8



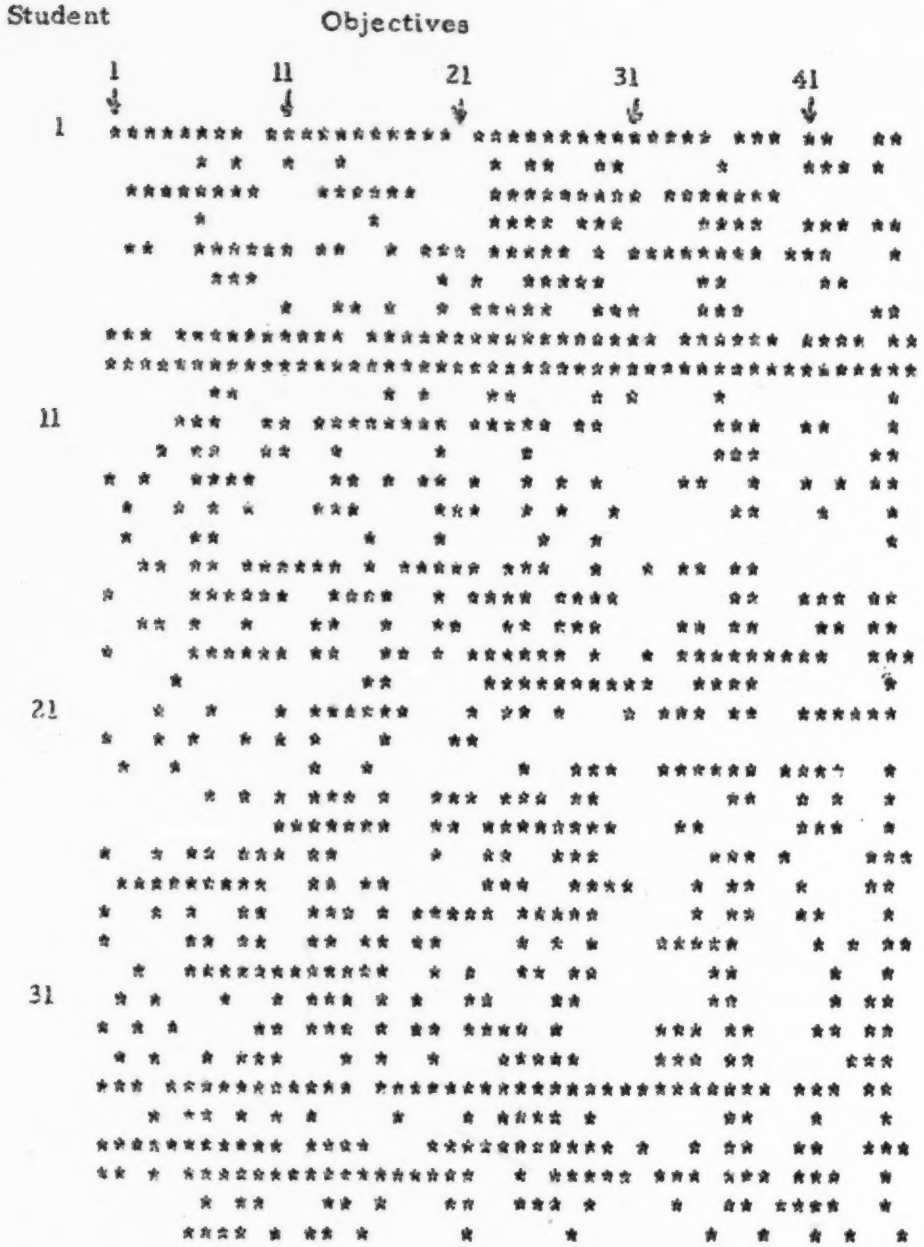
School B-Grade 8



School C-Grade 8

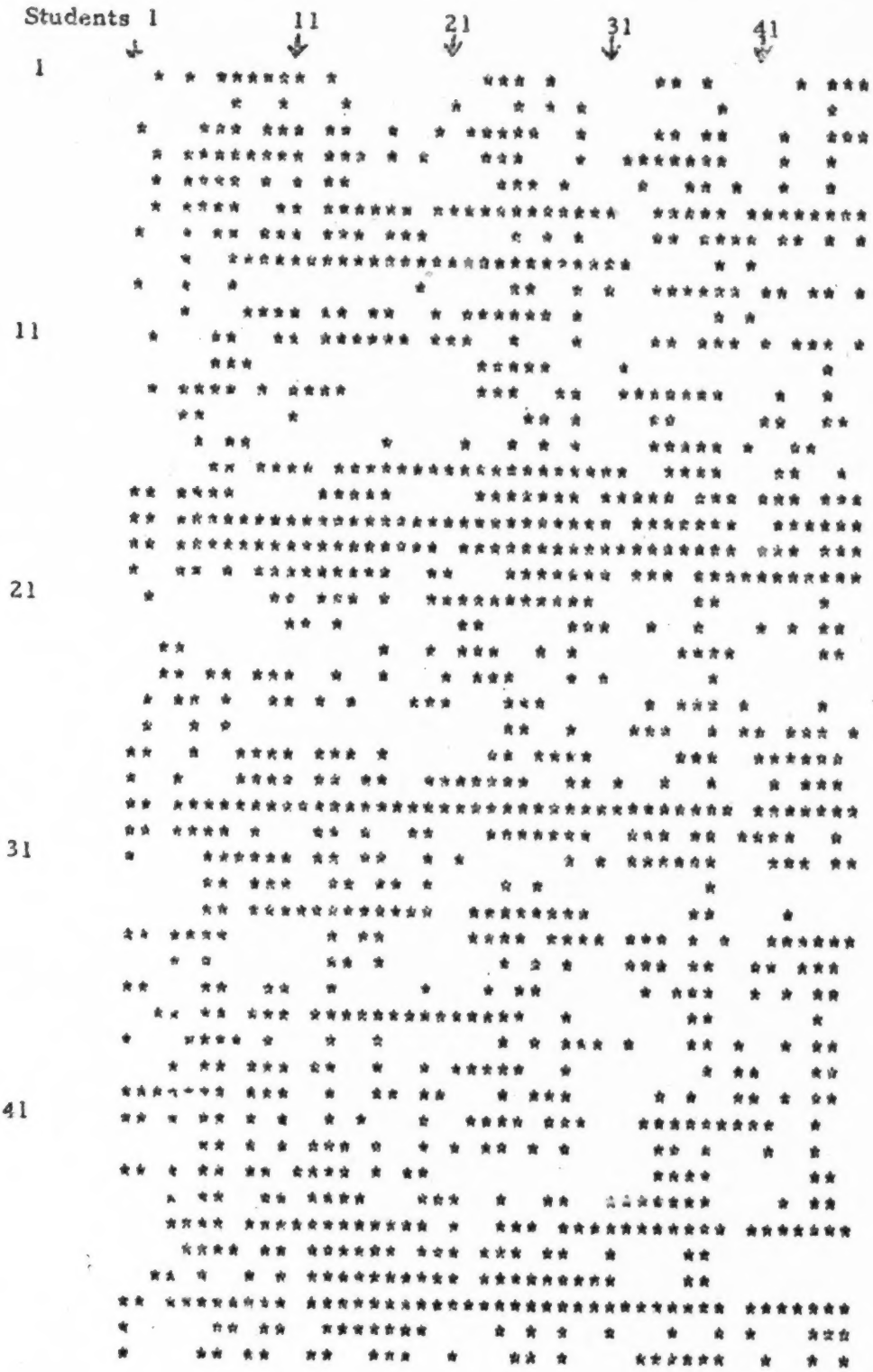


SCHOOL A - Grade 5

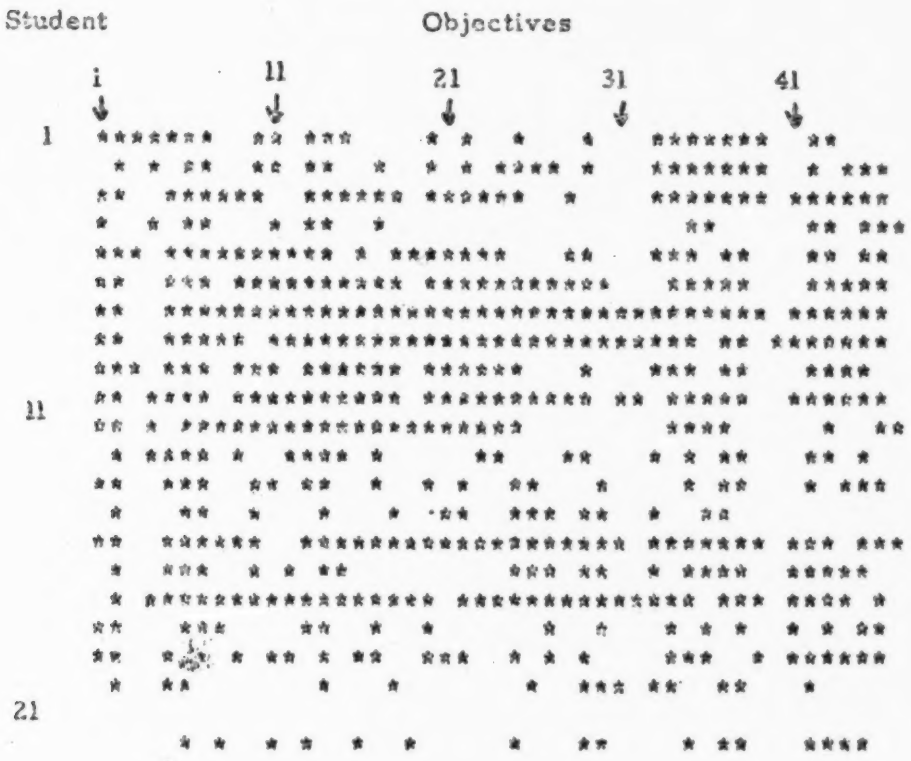


School B

Objectives



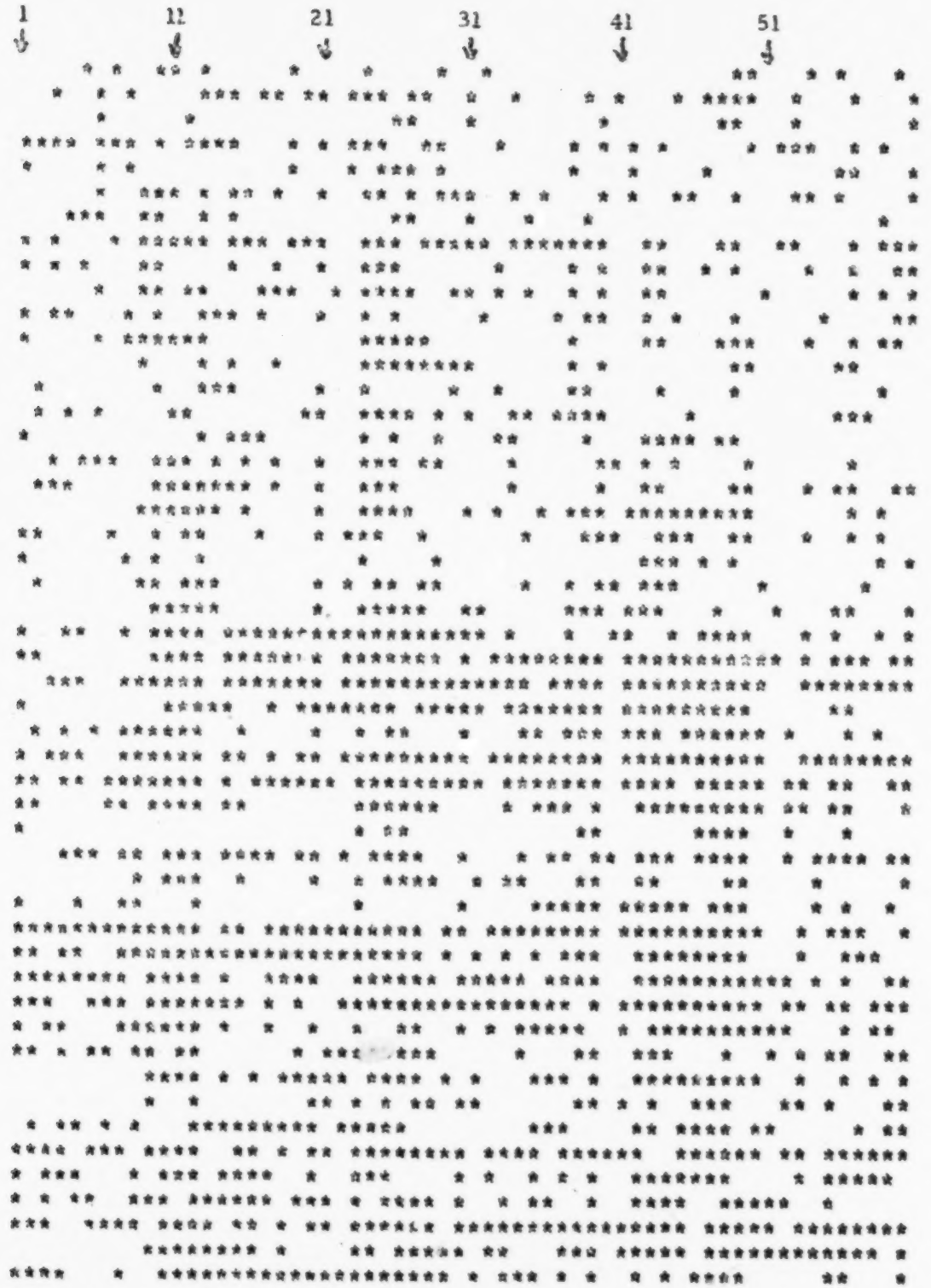
SCHOOL C - Grade 5



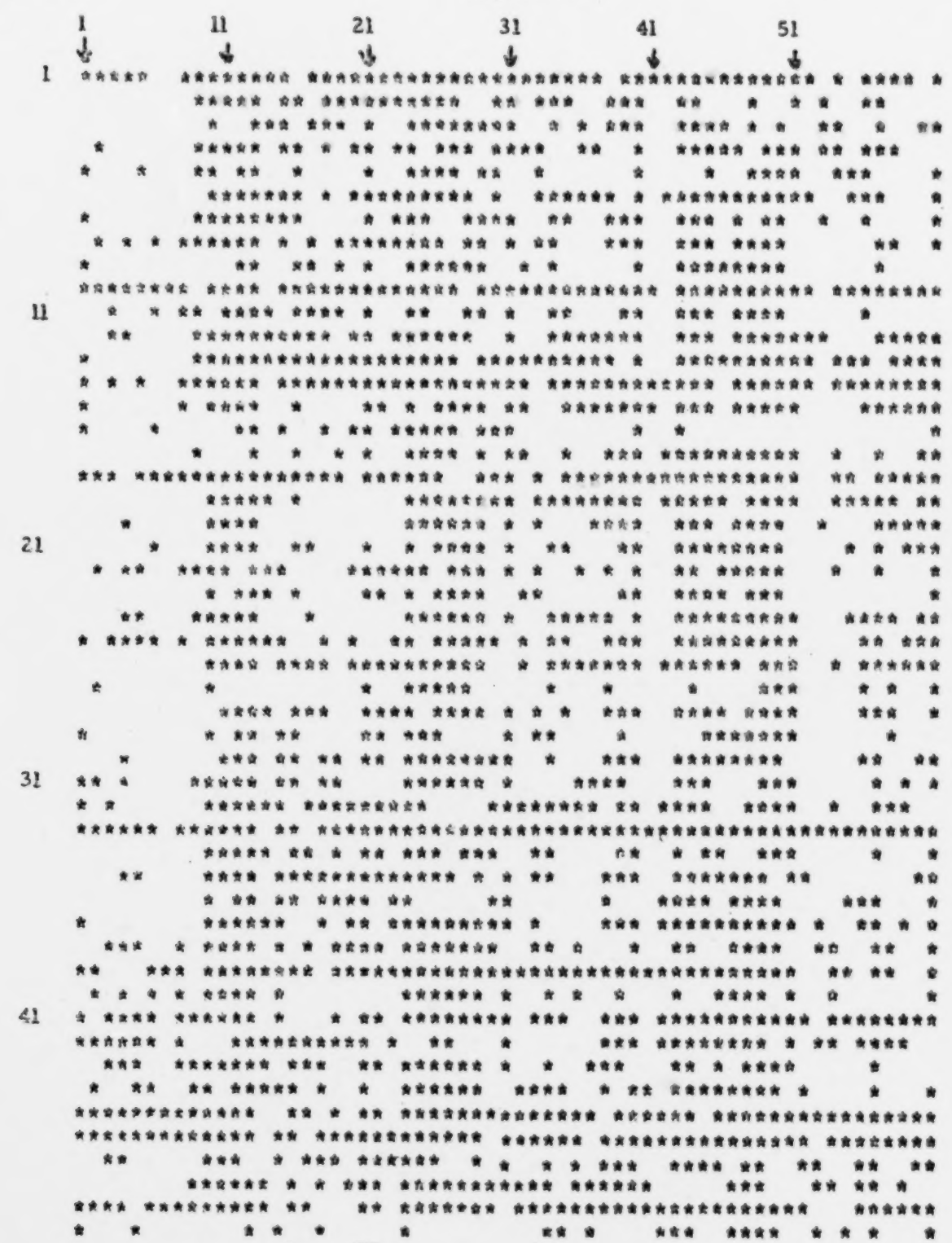
TEACHER :A- Grade 8

Student

Objectives



TEACHER B - Grade 8



Part 111 Teacher Attitudes Toward Criterion Referenced Tests

After one full year (1971-72) and the first semester of 1972-73, teachers who had participated in the criterion referenced testing program were administered in early February, a brief attitude inventory. All elementary teachers in grades 3 through 6 and junior high (grades 7-9) mathematics teachers comprised the potential respondents.

The inventory contained fourteen questions each answered by checking one of five choices from Strongly Agree to Strongly Disagree. Emphasis was placed on questions which dealt with criterion-referenced as opposed to non-referenced tests. Since individual item responses were desired, and time was extremely short, no effort was made to secure reliability or validity data other than face validity which the reader can judge for himself since the fourteen items comprising the inventory appear below.

1. I found Criterion Reference Tests more difficult to administer than standardized tests, such as the Iowa Tests of Basic Skills.

2. The language level of the Criterion Reference Test is appropriate for the grade level at which I teach.

3. The content of the Criterion Reference Test is more adaptable to local needs, than other standardized tests.

4. A Criterion Reference Test helps me to diagnois skill difficulties better than other standardized Tests.

5. The individual diagnostic printout of the Criterion Reference Test is more useful than the item analysis on standardized tests such as the Iowa Tests of Basic Skills.

6. The group diagnostic printout of the Criterion Reference Test is more useful than the item analysis on other standardized tests.

7. The content of tests such as the Iowa Tests of Basic Skills is more appropriate than the Criterion Reference Test to test achievement of students.

8. A Criterion Reference Test in other subject areas would be useful in improving the educational process.

9. The Criterion Reference Test can be used for placement purposes.

10. The Criterion Reference Test can be used at the end of the school year to assess student achievement.

11. When the Criterion Reference Test is being used for placement purposes at the beginning of the school year, the grade level form of the Criterion Reference Test should be the same as the students grade level.

12. The lack of norms for the Criterion Reference Test limits its usefulness.

13. Criterion Reference Tests relate closely to the districts curriculum objectives.

14. Criterion Reference Tests help teachers to achieve the teaching goals of our school district.

The following table contains the results obtained from administering the criterion referenced test to the 210 potential respondents of which 142, or 67.6 percent, responded. The responses are listed by item number referenced to the preceeding list of items. Three columns of data are provided for each item, elementary, junior high, and total. Each line contains frequency and percentage for the responses as well as means and standard deviations.

As might be expected the elementary teachers, for whom mathematics is only one of many subjects, tend not to have reached any group decision as to the value of criterion referenced tests. Approximately equal numbers responded Agree, Disagree, and Not Certain with the exception of items 2 and 11 where a significant number also responded Strongly Disagree and items 9, 10, and 13 were a greater proportion responded Agree. Basically, however, the results are such that no meaningful conclusions can be reached. It is suggested that the elementary teachers need more time and involvement(inservice training, etc.) before a more obvious attitudinal pattern will become apparent. Since baseline data is not available, no statistical judgement can be reached as to whether or not the results shown differ from what existed prior to criterion referenced testing.

The junior high teachers, all specializing in mathematics, tended to have reached more definite conclusions, although even here some obvious minority opinions are apparent. Nearly all of the teachers agreed that appropriate language was used in the tests(item 2). Two thirds of the teachers responded "favorably" to usage of criterion referenced tests in place of non-referenced tests such as the Iowa Tests of Basic Skills (items 1, 3, 4, 5, and 6). A similar proportion responded "favorably" to usage of criterion referenced tests for placement (item 9) and assessment (item 10) purposes. Finally approximately half of the respondents marked "favorable" responses (with another quarter checking Not Certain) to usage of criterion reference tests in other subject areas (item 8) and relationship to the district's instructional program (item 13 and 14).

The remaining three items reveal that, even among teachers who responded "favorably" to the usage of criterion referenced tests, there still lingers some support for non-referenced tests. For instance, nearly half of the teachers (with another quarter Not Certain) felt that the content of the Iowa Tests were more appropriate than the criterion-referenced tests for testing achievement (item 7) or that the lack of norms limits the usefulness of the criterion-referenced tests (item 12).

Data Pertaining to Responses of Elementary and Junior High School Teachers to the Desirability of Criterion-Referenced Tests

ITEM NUMBER	Group	SA*	A*	NC*	D*	SD*	$\bar{X}$	SD
1	Elem	15/11.9	28/22.2	33/26.2	38/30.2	12/ 9.5	2.97	1.18
	J.H.	0/0	3/18.8	3/18.8	8/50.0	2/12.5	2.44	0.96
	Total	15/10.6	31/21.8	36/25.4	46/32.4	14/ 9.9	2.91	1.17
2	E	3/ 2.4	31/24.6	28/22.2	25/19.8	39/31.0	2.48	1.23
	JH	3/18.8	12/75.0	0/0	1/ 6.3	0/0	4.06	0.68
	T	6/ 4.2	43/30.3	28/19.7	26/18.3	39/27.5	2.65	1.28
3	E	10/ 7.9	42/33.3	27/21.4	29/23.0	17/13.5	2.99	1.20
	JH	5/31.3	5/31.3	3/18.8	3/18.8	0/0	3.75	1.13
	T	15/10.6	47/33.1	30/21.1	32/22.5	17/12.0	3.08	1.21
4	E	7/ 5.6	46/36.5	26/20.6	27/21.4	20/15.9	2.93	1.19
	JH	2/12.5	8/50.0	4/25.0	2/12.5	0/0	3.63	0.89
	T	9/ 6.3	54/38.0	30/21.1	29/20.4	20/14.1	3.04	1.18
5	E	8/ 6.3	47/37.3	32/25.4	25/19.8	14/11.1	3.08	1.13
	JH	3/18.8	7/43.8	4/25.0	2/12.5	0/0	3.69	0.95
	T	11/ 7.7	54/38.0	36/25.4	27/19.0	14/ 9.9	3.15	1.12
6	E	2/ 1.6	39/31.0	44/34.9	25/19.8	16/12.7	2.89	1.04
	JH	2/12.5	8/50.0	4/25.0	2/12.5	0/0	3.63	0.89
	T	4/ 2.8	47/33.1	48/33.8	27/19.0	16/11.3	2.97	1.04
7	E	5/ 4.0	36/28.6	33/26.2	40/31.7	12/ 9.5	2.86	1.06
	JH	1/ 6.3	2/12.5	6/37.5	1/ 6.3	6/37.5	2.44	1.31
	T	6/ 4.2	38/26.8	39/27.5	41/28.9	18/12.7	2.81	1.10
8	E	8/ 6.3	35/27.8	46/36.5	27/21.4	10/ 7.9	3.03	1.03
	JH	1/ 6.3	6/37.5	6/37.5	1/ 6.3	2/12.5	3.19	1.11
	T	9/ 6.3	41/28.9	52/36.6	28/19.7	12/ 8.5	3.05	1.04
9	E	3/ 2.4	55/44.0	28/22.4	24/19.2	13/10.4	3.10	1.08
	JH	2/13.3	7/46.7	1/ 6.7	4/26.7	1/ 6.7	3.33	1.23
	T	5/ 3.5	62/44.0	29/20.6	28/19.9	14/ 9.9	3.12	1.09
10	E	4/ 3.2	54/43.2	26/20.8	25/20.0	13/10.4	3.10	1.10
	JH	1/ 6.7	10/66.7	2/13.3	1/ 6.7	1/ 6.7	3.60	0.99
	T	5/ 3.5	64/45.4	28/19.9	26/18.4	14/ 9.9	3.15	1.09
11	E	5/ 4.0	28/22.4	34/27.2	30/24.0	30/23.2	2.59	1.18
	JH	0/0	5/33.3	3/20.0	4/26.7	3/20.0	2.67	1.18
	T	5/ 3.5	33/23.4	37/26.2	34/24.1	32/22.7	2.60	1.18
12	E	9/ 7.2	35/28.0	40/32.0	33/26.4	6/ 4.8	3.06	1.02
	JH	0/0	7/46.7	4/26.7	3/20.0	1/ 6.7	3.13	0.99
	T	9/ 6.4	42/29.8	44/31.2	36/25.5	7/ 5.0	3.07	1.01
13	E	4/ 3.2	53/42.4	38/30.4	21/16.8	7/ 5.6	3.22	0.96
	JH	4/26.7	5/33.3	3/20.0	3/20.0	0/0	3.67	1.11
	T	8/ 5.7	58/41.1	41/29.1	24/17.0	7/ 5.0	3.27	0.98
14	E	4/ 3.2	42/33.6	30/24.0	28/22.4	11/ 8.8	3.00	1.02
	JH	2/13.3	5/33.3	4/26.7	3/20.0	1/ 6.7	3.27	1.16
	T	6/ 4.3	47/33.3	43/30.5	31/22.0	12/ 8.5	3.03	1.04

* Data presented as frequency/percentage

Conclusions

The authors view criterion referenced testing as a meaningful contribution to education, not as a panacea to all the ills of educational testing and measurement. It is time that more emphasis be placed on implementing this technique into practice. This paper provides some guidelines for so doing as well as pitfalls to avoid. In addition, this study indicates some of the unexpected byproducts of criterion referenced testing which, in themselves, justify the application of this technique.

The following are the adoption experiences of the school district with regard to criterion reference testing.

a. The language level on some of the lower level tests had to be redone.

b. The level 3 and 4 tests should be done with large type and large simple diagrams.

c. Teachers on the average favor criterion reference tests over national normed achievement tests.

d. At first some teachers feared the tests might be used for teacher evaluation.

e. Teachers favor both the individual student printout and the group printout.

f. The group printout groups students according to common skill deficiencies, and consequently it is a good indicator for remedial work and placement purposes.

g. The tests have showed that there is a different emphasis on skill objectives within schools and grade levels in our school district.

h. The tests provide teachers with a method for determining learner needs and accomplishments and offers a basis for measuring instructional accountability, which is extremely variable in parent-teacher and student conferences.

i. The test can be given at any time during the year for placement purposes, which is a problem in our district because of the many federally affiliated students.

j. The criterion reference test has projected certain weaknesses in our curriculum which we are able to deal with in a positive manner.

k. Very few students had trouble marking their responses on the mark sense computer cards that were used for scoring on each of the grade levels.

l. Test results tend to indicate that students were more concerned about doing well on the criterion reference tests than on various types of achievement tests. Simply, because they felt that their results on the tests would not require them to repeat previously learned skills.

m. The "branching effect" now needs to be implemented so that batteries of sub tests can be developed to further pinpoint the deficiencies within some of the larger skill areas.

n. The tests should not be given in one continuous sitting, about one hour is maximum time if you expect to get reliable test results.

o. All tests should be validated on a district wide basis with an outside validator in charge of the validation effort, who has the necessary background and expertise.

p. The district is now in the process of exploring the possibility of the concept of criterion reference testing in other subject areas such as reading and science.

q. The criterion reference test is the only type of test that a school district can use to determine if it is working toward its curriculum goals. Results to date have been encouraging and helpful

5	1	B	FINITE SET	321-100
5	1	E	INFINITE SET	321-100
5	1	A	DISJOINT SET	321-105
5	1	B	UNION	321-105
5	1	C	INTERSECTION	321-105
5	1	CD	VENN DIAGRAM	321-110
5	1	AB	REPLACE. & SOL. SETS	321-115
5	1	CED	ROUNDING NUMBERS	321-120
5	1	DAB	PLACE VALUE	321-125
5	1	CDE	PLACE VALUE	321-130
5	1	BDC	MULTIPLES OF A NUMBER	321-250
5	1	ADD	PRIME NUMBERS	321-253
5	1	BC	DIST. PROP. OF ADD. OVER X	321-256
5	1	BCA	2 DIGIT MULTIPLIERS	321-259
5	1	CDB	DIV. BY MULTIPLES OF 10	321-262
5	1	EEC	STORY PROBLEMS,DIVISION	321-265
5	1	CD	LONG DIV.,1 DIGIT DIVISOR	321-271
5	1	BAB	2 DIGIT DIVISOR	321-274
5	1	ADC	ZERO IN THE QUOTIENT	321-277
5	1	BD	FINDING AVERAGES	321-280
5	1	CDE	ADDITION,LIKE DENOMINATORS	321-300
5	1	BA	LEAST COMMON DENOMINATOR	321-305
5	2	CDD	ADDITION,UNLIKE FRACTIONS	321-310
5	2	EB	ADDITION OF MIXED NUMBERS	321-315
5	2	AC	SUBTRACTION,LIKE FRACTIONS	321-320
5	2	BD	SUBT.,UNLIKE FRACTIONS	321-325
5	2	CE	SUBTRACTION,MIXED NUMBERS	321-330
5	2	AB	(FRACTION)X(WHOLE NUMBER)	321-335
5	2	CA	MULTIPLY 2 FRACTIONS	321-340
5	2	EA	(WHOLE NUMBER)X(MIXED NO.)	321-345
5	2	BE	MULTIPLY 2 MIXED NUMBERS	321-350
5	2	CDE	PROB. SOLV.,FRACTIONS,+-X	321-355
5	2	ABB	PLACE VALUE,DECIMAL SYSTEM	321-400
5	2	CD	FRACTIONAL NAMES,DECIMALS	321-405
5	2	EA	FRACTIONAL NAMES,DECIMALS	321-405
5	2	CH	DECIMALS-LESS,GREATER THAN	321-410
5	2	ED	ADDITION OF DECIMALS	321-415
5	2	DE	SUBTRACTION OF DECIMALS	321-415
5	2	ACA	MULTIPLICATION OF DECIMALS	321-420
5	2	CAE	PROB. SOLV.,RATES & RATIO	321-500
5	2	BE	SCALE DRAWING	321-505
5	2	CD	PARTS OF A CIRCLE	321-600
5	2	BD	MEASUREMENT OF ANGLES	321-605
5	2	EDC	KINDS OF ANGLES	321-610
5	3	BA	TRIANGLES	321-615
5	3	CE	PERIMETER OF POLYGONS	321-620
5	3	D	CIRCUMFERENCE OF CIRCLES	321-625

P	1	BC	OTHER BASES TO BASE 10	441-005
P	1	AC	BASE 10 TO OTHER BASES	441-010
P	1	DE	ADDITION IN BASES	441-015
P	1	CB	MULTIPLICATION IN BASES	441-020
P	1	BD	SUBTRACTION IN BASES	441-025
P	1	A	BASE 2 TO BASE 10	441-030
P	1	E	BASE 10 TO BASE 2	441-030
P	1	DD	BASE 2 ADDITION & SUBTR.	441-030
P	1	CC	SET NOTATION	441-050
P	1	BC	ROSTER	441-050
P	1	DE	FINITE AND INFINITE	441-050
P	1	AE	GROUP SYMBOLS	441-055
P	1	BB	GROUP SYMBOLS	441-055
P	1	DD	GROUP SYMBOLS	441-055
P	1	C	COMMUTATIVE PROP. ADDITION	441-060
P	1	B	DISTRIBUTIVE PROPERTY	441-060
P	1	E	IDENTITY PROPERTY OF MULT.	441-060
P	1	C	COMMUTATIVE PROP. OF MULT.	441-060
P	1	E	IDENTITY OF ADDITION	441-060
P	1	A	ASSOCIATIVE PROP. OF ADD.	441-060
P	1	D	MULTIPLICATION PROP. OF 0	441-060
P	1	D	CLOSURE PROPERTY	441-060
P	1	A	ASSOCIATIVE PROP. OF ADD.	441-060
P	1	DE	INTEGERS, GRAPHING	441-085
P	1	EC	ADDITION, GRAPHING	441-085
P	1	AB	INTEGER ADDITION	441-085
P	1	BE	INTEGER SUBTRACTION	441-090
P	1	DC	INTEGER MULTIPLICATION	441-095
P	1	AD	INTEGER DIVISION	441-095
P	1	C	OPEN SENTENCES	441-105
P	1	A	GREATER THAN	441-105
P	1	B	LESS THAN	441-105
P	2	DA	EQUATIONS, ADDITION PROP.	441-110
P	2	EC	EQUATIONS, MULTIPLICATION	441-115
P	2	CD	EQUATIONS, 1 OR MORE PROP.	441-120
P	2	B	OPEN SENTENCES	441-125
P	2	C	OPEN SENTENCES	441-125
P	2	CD	OPEN SENTENCES	441-125
P	2	BB	PRIME FACTORS	441-155
P	2	AC	GREATEST COMMON FACTOR	441-160
P	2	D	RELATIVELY PRIME NUMBERS	441-160
P	2	BA	LEAST COMMON MULTIPLE	441-165
P	2	DB	ADDITION OF RATIONAL NOS.	441-180
P	2	CE	SUBTR. OF RATIONAL NOS.	441-180
P	2	DF	MULT. OF RATIONAL NOS.	441-180
P	2	EC	DIVISION OF RATIONAL NOS.	441-185
P	2	E	DECIMAL ADDITION	441-190
P	2	D	DECIMAL SUBTRACTION	441-190
P	2	CB	DECIMAL MULTIPLICATION	441-190
P	2	CC	DECIMAL DIVISION	441-195
P	2	C	RATIO TO DECIMAL	441-200
P	2	D	DECIMAL TO RATIO	441-200
P	2	A	ORDER	441-205
P	2	B	ORDER	441-205
P	2	C	RATIO	441-210
P	2	E	RATE	441-210
P	2	BF	PROPORTION	441-210
P	2	CAB	PERCENT PROBLEMS	441-215
P	2	BL	PERCENT WORD PROBLEMS	441-220
P	2	CCB	NON-MET. GEOMETRY	441-225
P	2	DE	GEOMETRIC CONSTRUCTION	441-235
P	2	C	RELATIVE ERROR	441-255